

Robust Group-Level Inference in Neuroimaging Genetic Studies

Virgile Fritsch^{*†}, Benoit Da Mota^{*†}, Gaël Varoquaux^{*†}, Vincent Frouin[†], Eva Loth[‡], Jean-Baptiste Poline^{†§}
and Bertrand Thirion^{*†}

^{*}*Parietal Team, INRIA Saclay-Île-de-France, Saclay, France*

Email: virgile.fritsch@inria.fr

[†]*CEA, DSV, I2BM, Neurospin bât 145, 91191 Gif-Sur-Yvette, France*

[‡]*Department of Forensic and Neurodevelopmental Sciences, Institute of Psychiatry, King's College London
London, United Kingdom*

[§]*Henry H. Wheeler Jr. Brain Imaging Center, University of California at Berkeley*

Abstract—Gene-neuroimaging studies involve high-dimensional data that have a complex statistical structure and that are likely to be contaminated with outliers. Robust, outlier-resistant methods are an alternative to prior outliers removal, which is a difficult task under high-dimensional unsupervised settings. In this work, we consider robust regression and its application to neuroimaging through an example gene-neuroimaging study on a large cohort of 300 subjects. We use randomized brain parcellation to sample a set of adapted low-dimensional spatial models to analyse the data. We combine this approach with robust regression in an analysis method that we show is outperforming state-of-the-art neuroimaging analysis methods.

Keywords—Robust regression, fMRI, neuroimaging;

I. INTRODUCTION

The statistical analysis of neuroimaging data is challenging since they are composed of multiple correlated descriptors (the images' voxels) the number of which is much larger than the number of observations. These data are observed in the presence of a complex structured noise. Subject performance, image acquisition, and data preprocessing are additional sources of variability that furthermore often lead to the presence of outliers into the datasets. These can cause dramatic drops in the performance of analysis methods. As the high-dimensional context prevents manual data screening, some outlier detection methods have to be used to provide a statistical control on subjects inclusion [1]. Yet, it remains unclear whether or not outliers should be removed, and, if so, what tolerance to choose. Alternatively, several outlier-resistant methods has been proposed for statistical inference in neuroimaging, although they are still not widely used. Beyond outlier-resistance, such robust methods seem better adapted to real world data since they also compensate for inexact hypotheses (e.g. data normality, homogeneous dataset). Wager [2] first showed that using *robust regression* (RLM) resulted in sensitivity improvements in group studies as compared to the use of *Ordinary Least Squares* (OLS) regression. Penny [3] and Woolrich [4] separate regular data from outliers with Bayesian mixture models. These studies however involve less than 20 subjects.

Here, we extend the work of [2]. We investigate the application of robust regression to larger cohorts and go beyond massively univariate testing by using the spatial structure of the data through parcel-based analyses. We consider a robust regression criterion that minimizes the sum of a convex function of the model residuals [5]. As the corresponding robust testing is only valid under specific conditions that are difficult to verify in practice, we use simulations to validate its use under our particular settings. Then, we apply the subsequent robust regression framework to a gene-neuroimaging study in (i) a voxel-wise analysis, (ii) independent parcel-based analyses, and (iii) a randomized-parcellations based analysis following the approach of [6]: parcel-based analyses are conducted using different brain parcellations and a consensus is then made so that the results are not parcellation-dependent while benefiting from the spatial structure underlying the images. We compare each procedure with its standard (non-robust) version, and show the sensitivity improvement that comes with robust regression. Randomized parcellations based inference with robust regression outperforms all existing methods in terms of sensitivity. Increased sensitivity is particularly vital for studies examining brain-behaviour relationships or gene-neuroimaging studies, which are sensitive to outliers and may compare groups with uneven cell sizes, for example due minor allele frequencies of common genetic variants.

II. TECHNICAL BACKGROUND

A. Huber's robust regression

Let \mathbf{y} be a variable corresponding to n observations, \mathbf{X} be a $n \times p$ matrix of p variables describing the same n observations, and ϵ be some noise. Considering a linear model $\mathbf{y} = \mathbf{X}\beta + \epsilon$, *robust regression* (in the sense of Huber [5]) intends to provide an estimate $\hat{\beta}$ of β by solving:

$$\hat{\beta}_{\text{RLM}} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^n \rho \left(\frac{y_i - \sum_j x_{ij} \beta_j}{\sigma} \right),$$

where ρ is a positive weighting function that dampens the influence of potential outlier values on the estimation. This

minimization problem is usually solved with the *Iteratively Reweighted Least Squares* algorithm presented below (Algorithm 1). In the sequel, we use Huber's weighting function with its default parameter $k = 1.345$: $\rho(x) = \frac{1}{2}x^2$ if $|x| \leq k$, $\rho(x) = k|x| - \frac{1}{2}k^2$ if $|x| > k$.

Algorithm 1 Iteratively Reweighted Least Squares

Require: $\mathbf{X}, \mathbf{y}, \rho$.

Init: $\epsilon = 10^{-8}$, $\mathbf{W}_{\text{old}} = \mathbf{I}$, $h(p)$ a normalization factor.

define function $\chi : x \mapsto x\rho'(x) - \rho(x)$

$\beta \leftarrow (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$

$\sigma^2 \leftarrow \text{Var}[y_i - \sum x_{ij}\beta_j]$

$\sigma^2 \leftarrow \frac{1}{n h(p)} \sum \chi\left(\frac{y_i - \sum x_{ij}\beta_j}{\sigma}\right) \sigma^2$

$\mathbf{W} \leftarrow \rho'\left(\frac{y_i - \sum x_{ij}\beta_j}{\sigma}\right)$

while $\|\mathbf{W}_{\text{old}} - \mathbf{W}\|_\infty > \epsilon$ **do**

$\mathbf{W}_{\text{old}} \leftarrow \mathbf{W}$

$\sigma^2 \leftarrow \frac{1}{n h(p)} \sum \chi\left(\frac{y_i - \sum x_{ij}\beta_j}{\sigma}\right) \sigma^2$ (scale step)

$\mathbf{W} \leftarrow \left(\frac{\rho'(y_i - \sum x_{ij}\beta_j/\sigma)}{y_i - \sum x_{ij}\beta_j/\sigma}\right)_{i \in \{1..n\}}$ (reweighting)

Let $\hat{\tau}$ be the solution of $\mathbf{X}^\top \mathbf{W} \mathbf{X} \hat{\tau} = \mathbf{X}^\top \mathbf{W} \mathbf{y}$

$\beta \leftarrow \beta + \hat{\tau}$

end while

$\text{cov}(\hat{\beta}) = K \frac{[1/(n-p)] \sum \rho''(y_i - \sum x_{ij}\beta_j)^2}{(1/n) \sum \rho''(y_i - \sum x_{ij}\beta_j)} \mathbf{W}^{-1}$, where K is a correction factor that depends on ρ

Huber [5] proposed to adapt the standard F-test to robust regression by considering a robust unbiased estimate of $\text{cov}(\hat{\beta})$ (given at the end of Algorithm 1). Such an analytic testing procedure is however crucial to us as the IRLS algorithm costs too much to be considered with permutation testing. We dedicate [subsection III-A](#) to a validation of this testing procedure as it has never been done to our knowledge.

We use a Python implementation of robust regression available in the statsmodels¹ library, which we optimized for our application. The implementation strictly follows Huber's definition of the scale update step, which ensures the algorithm convergence when a convex weighting function is used [5].

B. Randomized Parcellation-Based Inference (RPBI)

Following the approach of [6], we perform randomized parcellation-based inference from a finite set of brain parcellations \mathcal{P} and a parcel-based thresholding function θ_t defined as:

$$\theta_t(v, P) = \mathbb{I}_{\{F(\Phi_P(v)) > 0.1/\text{card}(P)\}} \quad (1)$$

where $\Phi_P : V \rightarrow P$ is a mapping function that associates each voxel with a parcel from the parcellation P ($\forall v \in P^{(i)}, \Phi_P(v) = P^{(i)}$), and F returns the F -statistic

¹<http://statsmodels.sourceforge.net>

associated with a given parcel-based average signal for a pre-defined test. \mathbb{I}_A is the indicator function of the set A . Finally, the aggregating statistic at a voxel v , $\sum_{P \in \mathcal{P}} \theta_t(v, P)$, is considered and tested for significance via permutation testing, yielding a voxel-wise p-values map similar to a standard group analysis map. Family-wise error control is achieved by tabulating the maximal value across voxels in the permutation procedure. Details about the generation of parcellations follow.

RPBI maintains the advantages of parcel-based analyses (i.e. reduction of the number of neuroimaging features, less sensitivity to inter-subject misalignment, and explicit consideration of the tests dependence) while making the results independent to the choice of the parcellation. We show that the combination of RPBI and robust regression outperforms robust voxel-wise analysis and OLS-based RPBI in terms of sensitivity.

III. OUTLIER-RESISTANT RANDOMIZED PARCELLATIONS BASED INFERENCE

A. Simulations

We carry out an empirical validation of the testing procedure proposed by Huber for robust regression, and we compare it with standard regression. We use the following model to generate n observations $\{y_1, \dots, y_n\}$:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{a}_q\epsilon + \alpha(1_n - \mathbf{a}_q)\epsilon, \quad (2)$$

where \mathbf{X} is a random $(n \times r)$ design matrix, β is the $(r \times p)$ matrix of the model coefficients, $\epsilon \sim \mathcal{N}(0, \text{Id}_n)$ models a Gaussian noise, \mathbf{a}_q is a n -dimensional vector with coordinates drawn from a Bernoulli distribution $\mathcal{B}(1-q)$, and $\alpha > 1$ is a scalar. Thus, q is the expected number of outliers in the generated dataset, and α is a parameter that controls how much the outliers deviate from the regular model. We set α to 5, potentially yielding gross outliers.

1) *Control of the type I error rate:* We first investigate control over the rate of type I error under the null hypothesis. We set a column of β to 0 in the model (2), the estimated coefficients of which we further test for nullity. For various contamination rates q , we fit both a standard and a robust linear model to 10,000 datasets generated according to model 2 and perform each time the above-mentioned non-zero significance test on the estimated coefficients. We expect that OLS/RLM reports a significant effect at $P < 0.05$ uncorrected in exactly 5% cases.

2) *Statistical power (type II error rate):* We show that in the presence of outliers, the statistical power of the robust test is higher than that of the statistical power achieved by an F-test subsequent to an OLS fit. The simulation framework is the same as in the previous experiment, except that we do not set any column of β to 0, so we perform tests on a variable that is known to have an effect. We construct Receiver Operating Characteristic (ROC) curves for RLM

and OLS so as to measure their sensitivity at any specificity level.

B. Application to a gene-neuroimaging study

We then applied this procedure to a study examining gene x environment (GxE) interaction effects on fMRI BOLD activity to angry faces in a large sample of 392 subjects. The example responds to the increasing recognition in the gene-neuroimaging field that genetic effects on brain function (and behaviour) may often only be detected under certain environmental conditions. Consequently, compared to main effects models, tests of the GxE interaction term render the need for sensitive neuroimaging "endophenotypes" all the more pertinent. As in many gene-neuroimaging studies we employed an unbalanced design, comparing 65 minor allele carriers of a common *Single Nucleotide Polymorphism* (SNP) in the oxytocin receptor gene (rs2268494) to 327 major-allele homozygotes. Severe outliers due to motion or deformation artefacts as well as those detected using a multivariate outlier procedure covering the whole brain, were removed.

We construct brain parcellations using Ward's clustering [7] on the contrast images of random subsamples of 300 subjects out of 1892, as the cohort we are working with is actually extracted from a larger imaging database that contains 1892 subjects, only 392 of which have the genetic information necessary to our study. Each parcellation is used to convert the contrast images of the 392 selected subjects into neuroimaging features by averaging the voxels signal within each parcel. This yields variable representations of the data that take into account the spatial structure of the signal at the population level, since we benefited from additional subjects to build the parcellations.

In a first experiment, we compare the ability of standard and robust regression to uncover significant effects when used with low-dimensional representations of the data coming from parcels averages. 200 brain parcellations (from 100 to 2000 parcels by increment of 100) are considered. For a given number of parcels and a given bootstrap sampling we generated 10 random parcellations. For each corresponding set of features, we conduct two analyzes: one with OLS and the other with RLM. We report the number of significant effects found ($P < 0.1$ Bonferroni corrected for the number of parcels) divided by the number of parcels, which gives a measure of the methods sensitivity, since we are confident from the simulation that we control the method's specificity. In a second step, we applied RPBI to this gene-neuroimaging study. We generated 100 random parcellations with 1000 parcels each following the above description. RPBI was performed twice: the first time with a standard regression algorithm (RPBI_{OLS}), the second time embedding a robust regression algorithm (RPBI_{RLM}). 1000 permutations were performed to convert the counting statistic into p-values.

We also perform a voxel-wise Bonferroni-corrected analysis, using respectively standard and robust regression.

IV. RESULTS

A. Simulation

1) *Type I error*: The control of type I error obtained with the testing procedures associated with OLS and RLM is exact, as shown in Figure 1. This results hold whatever the number of observations involved in the simulation. We also obtained the same performance when confounding variables were included, and when multivariate tests were considered (i.e. several columns of the design matrix were associated with null coefficients and tested for a joint effect).

2) *Type II error*: The ROC curves presented in Figure 2 illustrate the ability of the testing procedures associated with OLS (resp. RLM) to detect a significant non-null effect under the presence of outliers. The latter potentially mislead OLS while RLM keeps a good accuracy. The curves may drop as more confounding variables are included in the experimental design, but the relative performance of both regression algorithms is preserved.

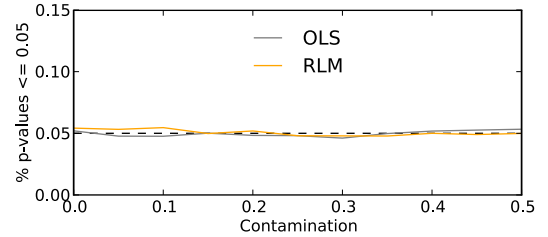


Figure 1. Proportion of type I errors for OLS and RLM, estimated on 10000 independent tests performed under a null hypothesis. The experimental design involves 300 observations ($n = 300$), 1 tested variable and 10 confounding variables.

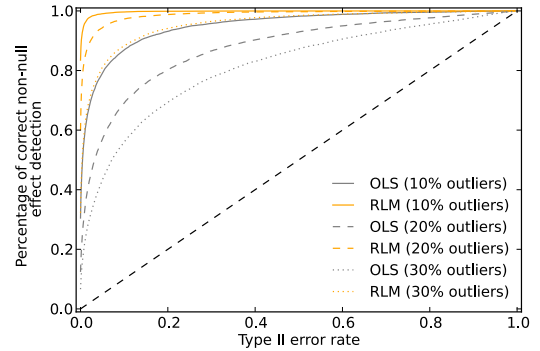


Figure 2. Accuracy of standard and robust regression algorithms under various amounts of contamination. Robust regression and its associated testing procedure always achieve a better compromise between type I and type II errors.

B. Gene-neuroimaging study

Figure 4 shows that robust regression always yields more significant activations than standard regression, for all number of parcels considered to reduce the data dimension.

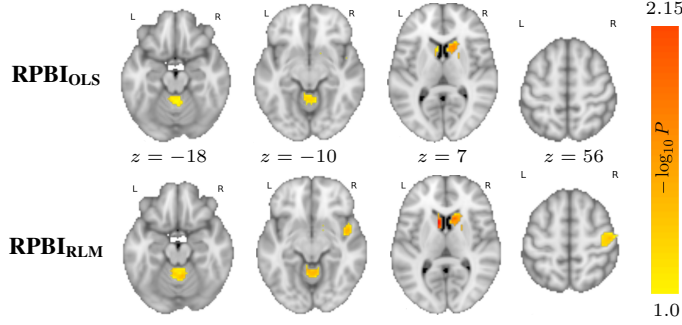


Figure 3. Voxel-level p-values maps given by RPBI_{OLS} and RPBI_{RLM} on our gene-neuroimaging study. Four brain regions are associated with a significant non-null effect according to the robust version of RPBI, while only two of them are reported by standard RPBI. The significant associations observed in the left and right ventral striatum (third column, $z = 7$) are particularly relevant to the study, as the ventral striatum plays a key role in the processing of positive and negative reward signals, including anger expressions.

Considering the definition of RPBI's aggregating statistic, this demonstrates that RPBI_{RLM} has a greater sensitivity than standard RPBI. As the proportion of reported significant activations stabilizes as soon as 500 parcels are used, **Figure 4** also suggests that the exact number of parcels does not have a strong impact if more than 500 parcels are considered (this was observed on other examples, not included in this work). This indicates that we can safely perform a parcellation-based analysis with a fixed number of 1000 parcels. Four brain locations were reported as significantly associated with a non-null effect when applying RPBI_{RLM} . Only two of them were reported by RPBI_{OLS} , as shown in **Figure 3**. Regarding voxel-wise analyses, only one (*resp.* four) voxel(s) located in the right (*resp.* right and left) ventral striatum passed the $P < 0.1$ Bonferroni-corrected threshold with OLS (*resp.* RLM), which illustrates the benefits of using parcellations on the present study. Randomizing the parcellations improves the results again by making them independent of a single brain representation, and robust inference improves again the sensitivity of the detections with (i) increased significance of the activations and (ii) a larger number of detections.

V. CONCLUSION

Gene-neuroimaging studies involve high-dimensional data that have a complex statistical structure and that are likely to be contaminated with outliers. Specific statistical procedures are therefore required to address this challenging problem. We have combined robust regression and Randomized Parcellation Based Inference in such a procedure that is sensitive, stable, and outlier-resistant in simulations. We demonstrated that robust regression can be used with an analytic testing procedure so that its embedding within RPBI is computationally affordable. We then showed that our method has an improved sensitivity by considering an application to a gene-neuroimaging study.

ACKNOWLEDGMENT

This work was supported by Digiteo grants (HiDiNim project and ICoGeN project), the ANR-10-BLAN-0128 grant, and the Microsoft Inria joint center grant *A-brain*. The data were acquired within the Imagen project. JBP was partly funded by the Imagen project, which receives research

funding from the E.U. Community's FP6, LSHM-CT-2007-037286. This manuscript reflects only the author's views and the Community is not liable for any use that may be made of the information contained therein.

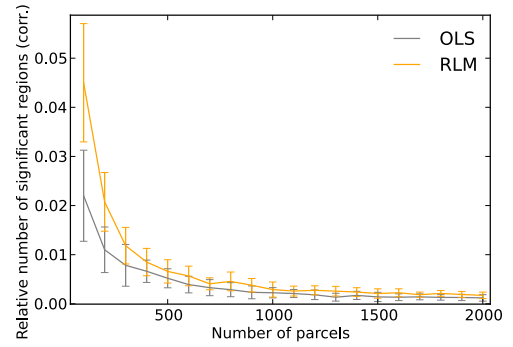


Figure 4. Percentage of parcels significantly associated with a non-null effect of the $\text{SNP} \times \text{Stressful Life Events (SLE)}$ interaction, according to standard and robust regression. The latter always uncover more significant associations, which makes it a good candidate for a combined use with Randomized Parcellations Based Inference.

REFERENCES

- [1] V. Fritsch, G. Varoquaux, B. Thyreau, J.-B. Poline, and B. Thirion, "Detecting outliers in high-dimensional neuroimaging datasets with robust covariance estimators," *Med Image Anal*, vol. 16, pp. 1359 – 1370, 2012.
- [2] T. D. Wager, M. C. Keller, S. C. Lacey, and J. Jonides, "Increased sensitivity in neuroimaging analyses using robust regression," *NeuroImage*, vol. 26, p. 99, 2005.
- [3] W. D. Penny, J. Kilner, and F. Blankenburg, "Robust bayesian general linear models," *Neuroimage*, vol. 36, p. 661, 2007.
- [4] M. Woolrich, "Robust group analysis using outlier inference," *Neuroimage*, vol. 41, p. 286, 2008.
- [5] P. J. Huber, *Robust Statistics*. John Wiley & Sons, Inc., 2005, ch. 7, p. 149.
- [6] G. Varoquaux, A. Gramfort, and B. Thirion, "Small-sample brain mapping: sparse recovery on spatially correlated designs with randomization and clustering," in *Int Conf on Mach Learn*, 2012.
- [7] J. Ward, "Hierarchical grouping to optimize an objective function," *J Amer Statist Assoc*, vol. 58, pp. 236–244, 1963.